(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 November 2004 (25.11.2004)

PCT

(10) International Publication Number
WO 2004/102858 A2

(51) International Patent Classification[7]: H04L

(21) International Application Number:
PCT/US2004/015030

(22) International Filing Date: 13 May 2004 (13.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/470,000 13 May 2003 (13.05.2003) US

(71) Applicant and
(72) Inventor: **COHEN, Hunter, C.** [US/US]; 16 Water Mill Heights Drive, Water Mill, NY 11976 (US).

(74) Agents: **BRUNELL, Norman, E.** et al.; Irell & Manella LLP, 1800 Avenue of the Stars, Suite 900, Los Angeles, CA 90067-4276 (US).

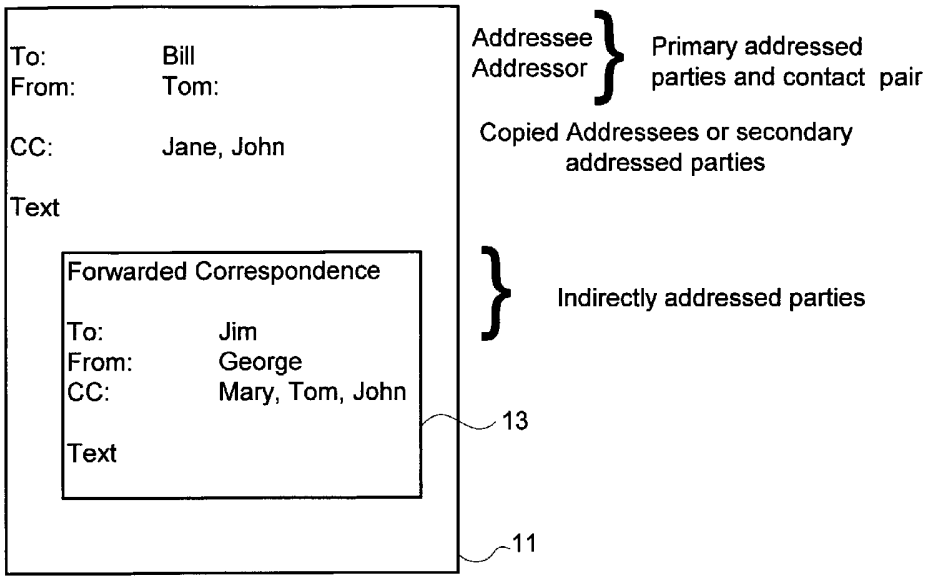
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for all designations*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

[Continued on next page]

(54) Title: DERIVING CONTACT INFORMATION FROM EMAILS



(57) Abstract: Correspondence, such as emails, is processed to develop a database of relationships between parties addressed on the correspondence including indirectly addressed parties such as those directly addressed in included, forwarded correspondence. The database may be used to determine the contact paths between users and addressed parties including the intermediary contacts required to complete contacts paths to selected addressed parties. Patterns of correspondence, including frequency and recency of correspondence may be detected and displayed. Statistically normal patterns of correspondence may be derived in order to determine if correspondence patterns for selected addressed parties deviate there from.

WO 2004/102858 A2

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DERIVING CONTACT INFORMATION FROM EMAILS

Background of the Invention

1. Related Applications:

This application claims the priority of US Provisional Application Serial No. 60/470,000 filed May 13, 2003.

2. Field of the Invention:

This invention is related to data mining and in particular, to deriving and using relations and patterns of relationships from collections of correspondence and the like, such as e-mails.

3. Description of the Prior Art:

We have all had the experience of meeting someone for the first time and quickly discovering that you are "connected" by an unexpected chain of acquaintances, often a short chain of only two or three people. In fact this occurrence is so common that we have a catch phrase response that most everybody uses "It's a small world", and even a play based on the phenomena, John Guare's "Six Degrees of Separation".

With the U.S. population just over 290 million and almost 6 billion more in the rest of the world, how can this "small world phenomena" be such a common occurrence, and is there a way to systematically employ it to our benefit?

What are needed are techniques for determining and using data to derive and exploit these chains of acquaintances.

Summary of the Invention

In a first aspect, a method for developing contact information from correspondence such as emails includes processing a set of correspondence to develop a database of relationships between addressed parties provided by one or more users, maintaining the database by further processing later received correspondence, and utilizing the database of relationships to provide relationship information between at least one of said users and the addressed parties.

A unique identification may be associated with each piece of correspondence and used to detect duplications of correspondence in order to more accurately determine a frequency of communication between addressed parties. The database may be maintained on a web based database of relationships in which addressed parties from a plurality of users are combined. Directly and indirectly addressed parties may be processed in correspondence to develop the database of relationships.

Connection paths between each of said users and at least some of the addressed parties may be displayed and additional addressed parties may be displayed upon selection of certain displayed addressed parties. Intermediate addressed parties, if any, between users and a selected addressed party may be visually displayed and/or prioritized together with the frequency of correspondence as well as the most recent correspondence between at least some of said addressed parties. The connection paths may be displayed, and/or prioritized in accordance with the closest, most recent, most frequent or some combination of recency, frequency and proximity of the correspondence between users and a selected addressed party.

Incoming correspondence may be sorted in accordance with the number of intermediate contacts, if any, identified in the database of relationships between users and the addressors of said incoming correspondence. Outgoing correspondence may be addressed to addressed parties in the database selected in accordance with the number of intermediate contacts, if any, between users and the addressed parties. Data related to the skills and experience of third parties may be processed to identify paths between users and third parties having selected skills and experience. Data related to the shopping experiences of third parties may be processed to identify paths between users and third parties having selected shopping experiences. The database of relationships may be analyzed in accordance with statistic norms to determine any deviations from such statistical norms of the correspondence pattern of selected addressed parties.

In another aspect, a method for deriving qualitative information related to addressed parties on correspondence such as emails includes processing a set of correspondence to develop a database of relationships between addressed parties, maintaining the database by further processing later received correspondence, and utilizing the database of relationships to determine patterns of correspondence for one or more of said addressed parties. Indirectly addressed parties on the correspondence may be processed to develop the database of relationships between directly and indirectly addressed parties.

Unique identification numbers may be associated with each piece of correspondence and used to detect duplications of correspondence in order to more accurately determine a frequency
of communication between said addressed parties. The database of relationships may be maintained on a network, such as the web, in which addressed parties from more than one user may be combined. The frequency of correspondence, and the most recent correspondence, in the database of relationships between addressed parties may be determined. Normal patterns of correspondence between addressed

parties may be derived to determine patterns of correspondence for a selected addressed party is consistent with the derived normal patterns.

In a still further aspect, a method for developing contact information from a user's correspondence such as emails, includes processing a collection of the user's correspondence to develop a database of relationships between said user and parties directly and indirectly addressed in said correspondence, maintaining the database by further processing later received correspondence, and utilizing the database of relationships to provide relationship information between the user and the addressed parties. A unique identification may be associated with each piece of correspondence and used to detect duplications of correspondence before maintaining the database in order to more accurately determine a frequency of communication between the user and the addressed parties. The database may be maintained on a web based database of relationships in which addressed parties from other sources may be combined. Connection paths between the user and at least some of the addressed parties may be displayed and additional addressed parties may also be displayed upon selection of certain displayed addressed parties.

Further displays may include intermediate addressed parties, if any, between the user and a selected addressed party, the frequency and most recent correspondence between the user and selected addressed parties while connection paths may be prioritized in accordance with the number of intermediate addressed parties, the most recent correspondence and/or the frequency of correspondence between said user and said pre-selected addressed party. Incoming correspondence may be sorted in accordance with the number of intermediate contacts while outgoing correspondence may be addressed to parties selected in accordance with the number of intermediate contacts.

Brief Description of the Drawings

Fig. 1 is an illustration of a correspondence document including primary and secondary directly addressed parties as well as a forwarded document including a series of parties indirectly addressed in the correspondence document.

Fig. 2 is a visualization of the various contact paths, and some of the contact path secondary information related to contact path direction, of the document shown in Fig. 1.

Fig. 3 is a top level, block diagram flow chart of the operation of the overall technique disclosed for creating and using a database of contacts collected from email records.

Fig. 4 is a block level flow chart of the relationship visualization aspects of the technique.

Fig. 5 is a display of a relationship tree illustrating the contacts for User A.

Fig. 6 is a block level flow chart of the referral path identification aspects of the technique.

Fig. 7 is a display of a selected referral path in the relationship tree of Fig. 5.

Fig. 8 is a block level flow chart of the SPAM filter.

Fig. 9 is a block level flow chart of the marketing tools aspects of the technique.

Fig. 10 is a block level flow chart of the skill and experience based path selection aspects of the technique.

Fig. 11 is a block level flow chart of the interface with third party software developers.

Fig. 12 is a block level flow chart of the shopper connection aspects of the present invention.

Fig. 13 is a block level flow chart of the mail scoring service aspects of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

Why the "small world phenomena" occurs in the first place we believe is a function of the following factors. The average person has a loose clique of friends and acquaintances that form based to a considerable extent upon happenstance, but strongly influenced by a number of less random factors such as an individuals job position and location, schools attended, schools children attend, financial status, hobbies, religious practices, commuting habits, stores frequented, participation in community activities, and the long list of other activities that comprise everyday life. The "circle of acquaintances" that make up these cliques appear typically to number from 200 to 400 individuals. Obviously there are exceptions to the rule, the recluse that knows only his mailman, or the town socialite who seems to know everyone, and the actual number depends on many circumstances. For convenience, an average number of 300 individuals in a circle of acquaintances will be used.

Almost by definition, the nature of these cliques causes many if not most of the members to share essentially the same acquaintances. Inevitably however, if an arbitrary member, let's call her Sally, carefully maps the relationships between all the people she socializes with, she will find that a small subset of her clique will know almost none of the other members except for those where Sally made the introduction. These friends that are members of Sally's clique solely by virtue of their relationship with Sally are usually strong links to other cliques and may be called "nexus contacts". There appear typically to be on the order of about 5 to 15 nexus contacts per clique, for this discussion, an average of 10 will be used. These nexus contacts, although linked to Sally's clique only by Sally, are typically strongly linked to one or more other cliques, also with about 300 individuals. These linked circles of acquaintances include multiple chains of acquaintances, as discussed above and may be used to identify potential contact paths between individual and may also be used to create actual contact paths, by for example referrals, between individuals.

The individuals within a clique are generally not randomly distributed throughout the general population, however, when we look at a similar size group of "linking" or "nexus" contacts, they are distributed throughout the general population in a surprisingly random pattern. Furthermore, when a small percentage of the population is represented, there is relatively little overlap in the membership between cliques that are connected by the nexus contacts. It is a consequence of this pattern of connection, that the number of individuals just a few handshakes away grows geometrically.

This geometric pattern of growth means, in the idealized case, that the average person is only six introductions away from over 300 million people. The idealized case assumes an average clique size of 300, each with 10 nexus individuals and no overlap in member constituents between cliques. The bottom line, if you are looking for an introduction to a specific person, there is a very good chance that they are within a few degrees of separation from you. The degrees of separation between two people in this context means the number of intermediary contacts needed to perform an introduction. For example, if Joe knows Sally and wants an introduction to Mary, one of Sally's friends, the degree of separation between Joe and Mary is one degree of separation because one intermediary, Sally, would be required to make an introduction or provide a referral between Joe and Mary.

A technique is disclosed for determining which introductions you need to get to a person you are trying to reach, using information related to addressed parties derived from correspondence, using emails as an example. A personal and private relationship tree is derived from a database of relationships which may be derived from some or all of the addresses of addressed parties included in emails sent or forwarded to you, and then, in a clear and actionable format, the possible contact paths, or paths of introduction, to the person you are trying to reach may be displayed and used. The technique need not be limited to email communication and is applicable for other types of correspondence where a record of the communicating parties may be made available electronically. Examples include phone records, as from telephone bills, instant messaging logs, or similar compendiums of contact data.

The term "Relationship Finder" refers to the techniques for automatically building a personal and private relationship tree and the tools to access this information.

The terms "Nexus Quotient" (or NQ) and "Estimated Nexus Quotient" (or ENQ) refer to two methods of providing a normalized measure of the extent of an individual's connections as evidenced by his or her communications history.

The term "World View" refers to an online subscription service that can be used to expand the reach of a user's database by enabling password protected access to the relationship trees of other subscribers in one or more predefined groups.

The term "Skills Registry" refers to an online service where individuals record their education, expertise, skills and experience, enabling users to search their relationship trees for introductions to people with specific qualifications.

The term "Referral Marketing Toolkit" refers to techniques allowing users to market products to their relationship tree through qualified referrals from people they know.

The term "SpamGate" refers to techniques for using knowledge of the addresses in a user's relationship tree to intelligently filter out unwanted bulk email solicitations, while insuring that all the messages they want get through.

The term "email scoring service" refers to a service that scores an email address based upon its observed frequency and pattern of communication as compared to some statistical norm. One of the possible uses for the email scoring service is to provide a predictive assessment of the likelihood that a particular address is being used for valid commerce versus dishonest use. That is, an email address may be scored to indicate that it has been involved in a normal pattern of communications for a reasonable length of time or it may be scored to indicate that it has been used in a pattern of communication, such as only for outward bound mailings, that is not indicative of a normal email address for an individual. This information may be

arrived at without regard to the identity of the email address holder and without regard to any specific individuals with whom communication has taken place.

Referral endorsement services refers to a service that can be integrated with retail commerce websites, auction websites, and other public websites with the purpose of providing website visitors a means to obtain website specific endorsements and or references from individuals they know or can reach indirectly.

The Email Relationship Finder may be provided as a "stand alone" software product or as a "plug-in" to Microsoft Outlook® and Outlook Express® or other email clients and may run on Microsoft Windows® 95, 98, 2000, NT and XP or other operating systems. In other embodiments, the Email Relationship Finder may work directly (either client-side or server-side) with POP3, MAPI, IMAP, and Hotmail or similar compliant online email account protocols.

The Email Relationship Finder may be used for extracting email or addressed party relationship pair information and also may serve as a user interface to the other services. The discovery of additional email stores, and the selection of logical locations to search for additional valid addresses, may be valuable steps in expanding the breadth and depth of a database of relationships. For instance, consider that in Microsoft Outlook, it would not be prudent to search the "inbox" or "deleted" folders since they will invariably contain "spam" from people with whom the user has no relationship. In an alternate embodiment, it is possible to optionally maintain separate lists to process, each with multiple folders to search, in the event users wish to maintain separate relationship trees, such as business, personal, school, etc. A given folder may reside on multiple lists. In still another embodiment, discovery and/or selection of folders may happen automatically and all emails could be analyzed without concern of pre-selection. In this embodiment, global information related to spam characteristics may optionally be employed to eliminate those communications from analysis.

Extraction, or parsing, of email addresses from all email headers and positional recognition of email addresses in text files, such as may be found in forwarded attachments, is an important step in the process. Extraction may be limited to the directly and indirectly addressed parties by for example extracting addresses following the "From:", "To:", and "Cc:" markers on the email correspondence being processed and as well as on forwarded emails attached thereto. The extraction process may optionally also extract secondary information, when present, related for example to the direction of the correspondence by extracting the email text labels attached to the email address and the date of communication (either sent date or received date). The email internet ID may also be extracted for use in preventing duplicate emails from being parsed.

The process may provide the automatic building and maintenance of databases of relationships, such as relationship tree databases, on a logical local drive that may optionally be user selectable, from all extracted email addresses and "screen names" automatically as part of the extraction/parsing functionality. In an alternate embodiment, separate relationship trees may be maintained matching the separate lists of grouped folders processed.

The user may have control over and may maintain preferences for his relationship tree with respect to database sharing and privacy in conjunction with the online services. In the alternate embodiment, the user may have control over, and may maintain preferences separately, for each relationship tree.

Optional embodiments may provide the user the ability to:

1) Maintain of a list of alternative (alias) email addresses that the user uses. All link searches may begin by default with these addresses.

2) Maintain lists of alias email addresses for their contacts so that all alias addresses may be automatically known to be the same contact when performing searches.

3) Maintain a global list, and individual lists, of email addresses to exclude from the relationship tree databases.

The data stored in the relationship tree databases may contain additional or secondary information, but for each instance of every email address pair extracted, the following information typically may be collected and stored:

The email addresses forming each "end" of the email pair.

The latest email communication date.

A pointer linking the email addresses that defines the contact pair relationship and direction of communication and the frequency of communication between the two addresses.

A unique original email ID# to prevent duplicate processing. This is collected for each message processed, not each pair.

In alternate embodiments, the relationship database may be cross referenced to other local, public, or private third party databases that are indexed by email address and contain relevant information that may be of interest either as a search term or a search result.

The following reporting options may also be made available:

Ability to list all email addresses alphabetically by the degree of separation or visa versa.

Ability to export email addresses to spreadsheets, with degrees of separation, or address books, with the category coded to show the source relationship tree name and degree of separation.

Ability to choose target email addresses with a list of alternates because many people have several email addresses.

Ability to maintain several lists, that the user can select or deselect, of email addresses to exclude from email chains.

Ability to choose up to how many degrees of separation to report.

Ability to change default maximum number of linkages to show.

Ability to choose date range to include based upon email received date.

Ability to list which email relationship trees to run search on.

Ability to override the default origin address and input a separate address to view chains between other individuals.

Report Display options may include:

View on screen a text based report of results.

View on screen a graphic report or display of results.

Write to word processing file.

Write to spreadsheet file (by degree of separation and for 1st degree or greater showing link addresses in successive columns).

Display/Hide date of email.

The World View may be available by subscription that allows users to share selected personal relationship tree databases via a centralized online database and to gain access to a larger universe of email address paths than they have individually. Access to the shared trees may be limited to the addresses on the direct path between addresses contained on the subscribers database and the target address. Therefore subscribers may only be shown email address information on paths that originated in their personal contact trees and end with the target address, i.e. the shared and personal relationship trees connect through a common email address. In other embodiments, users that share access may have full view of each others' information.

Each user optionally may maintain a list of email addresses that are to be excluded from the shared tree. Any time excluded addresses are encountered, those addresses, and any down-line addresses in those chains, may not transferred to the online database.

The Skills Registry may consist of two web based components that together allow introduction paths to people to be determined based upon the "target's" qualifications rather than knowledge of their email address.

The first component of the Skills Registry is a web based registry that may allow any individual, whether or not they are users of the email relationship finder, to enroll in the service and record their education, expertise, skills and experience on a secure and restricted database. The enrollee can revisit the site at any time to update or modify their profile. The profile is compiled by selecting from an extensive list (with optional temporal

qualifiers; such as when, how long) of job functions, job titles, company names, school degrees, schools attended, professional development programs, professional expertise, geographic information, family information, hobbies, interests, etc. Free form information may be the contact information, address, telephone, etc., and a non searchable file attachment, typically a resume, curriculum vitae, or portfolio. The amount of information provided is at the discretion of the enrollee. The enrollee must enter at least an email address. Each email address entered may receive a coded reply that may require a separate response before it is authorized in order to insure the validity of the address and its owner information. The enrollee may also enter the maximum distance in degrees of separation that a inquirer can be from the enrollee in order to have access to this information. The Profile information is used to generate search results. The free form information, if any, is provided to inquirers that find the enrollee as a result of a profile search. In either case, the information can be restricted so that it is only accessible to inquirers within the distance defined by the enrollee. As an incentive to enroll in the registry, registry users may be offered the option of learning their Estimated Nexus Quotient (ENQ) which is based largely upon the frequency and position that their email address appears in the global database of all users.

The second component allows World View users to search their relationship trees for introductions to people with specific qualifications.

The Referral Marketing Toolkit© allows users to market Email Relationship Finder and other select products. Once the software is installed, a popup window may periodically present an offer to promote the Email Relationship Finder product, and selected other tools, to all zero and one degree of separation email addresses, i.e. those addresses that have had direct contact with the user and need no intermediary introduction or need only one intermediary introduction. The offer may provide some form of compensation, such as cash for each unit sold to the first degree address holder, or as a prize based upon the most units sold by referral, or with earned MLM points that are good to redeem products. When a purchaser is referred by more than one source or more than one time, each referrer that provided the introduction prior to the purchase fractionally shares the credit. A "multi-level marketing" or MLM

version of this promotion plan allows credit to be awarded for "down line" sales as well.

If a user agrees to participate in the promotion, then the user can choose from a short list of pre-scripted promotional letters where a portion is user editable. The letter is from the registered user's email address and each copy is individually addressed to all zero and one degree email addresses in the users contact tree. When the user sends out promotions, the zero degree and one contact list is sent to a mail server that handles the outbound mailing for the user avoiding ISP bulk mail restriction issues, and at the same time, this facilitates tracking of referrals for reward purposes. Each promotion has a unique identifier and the list server will only send the first 3 of a given promotion to an individual. This avoids over mailing popular promotions from a large number of users. If the user does not participate in the promotion, they are asked again periodically. An option to turn off this prompting is available.

From time to time, active users may be offered to promote selected products using the same method and with various compensation or prizes.

Extended functionality may be available in which a special email composition tool may be provided for the user to market their own products.

Other embodiments of the referral marketing program may allow users to check off product types that they have interest in. When a user sends other users promotional letters, even through non-user intermediaries, they only go to those users that have interest in the types of products being marketed.

SpamGate is a spam filtering tool that in one embodiment works as follows:

1. SpamGate installation may add "quarantine" folders to the user's email client, such as: Inbox_Filtered; Inbox_FollowUP; Deleted_Spam; and Saved_By_Name. In addition, a toolbar may be added with selections such as Delete Content, Delete Email Address, Undelete, File As, Follow Up, and/or Auto File buttons.

2. When SpamGate is active, emails that arrive go through a "vetting" process to filter the incoming messages. The user first decides how many degrees of separation on their relationship tree to use when matching incoming email addresses with relationship tree addresses. The assumption is that spam will not be coming from email addresses that are part of acceptable correspondence. When a "From:" email address matches a relationship tree address, the email goes into a special inbox-filtered folder otherwise it goes to the normal inbox.

In one embodiment, as the users view email in their normal inbox, they have several options:

1. They can move the email to a folder set to process addresses into a relationship tree and therefore add the addresses to a vetted list.
2. They can move the email to a folder set only to add the addresses to a vetted list but not process addresses into a relationship tree.
3. They can move the emails to a folder set to not do anything or use the normal delete key and the addresses will be added to none of the lists.
4. They can use the Delete Email address button and the address will be moved to a list where all future emails from that address will be deleted automatically. In the event that the address already exists in the user's relationship tree, the user is asked if that address should be deleted from the tree as well. If the answer is yes, then those address occurrences and all their down-line chains are removed as well.
5. They can use the Delete Content button and whenever the same content of the message arrives, regardless of the sender, the message will be deleted automatically. A formula converts each message to a unique number to accomplish the required matching. After the Delete Content key is pressed, the email does not move until either the normal delete key or the Delete Email address key is pressed (allowing the content and address to be placed on automatic delete lists as well, if desired).
6. They can use the Follow Up button and the email will be moved to the "Inbox_FollowUp" folder. A popup window asks when to follow up. When the

follow up date and time is reached, if the email is still in the folder, it is automatically forwarded, from screen name Follow_Up, to the Inbox_filtered folder using the then current date and time and it is marked as unread.

7. They can use the File As button and the email will be moved to a subfolder of the Saved_By_Name folder. A popup window asks to name the subfolder as either the sender's email address, the sender's screen name, or some other name that the user specifies. If the user had previously processed an email from the same sender email address using the File As button, then the popup window does not appear and the email is simply moved to the same folder as the prior time.

8. Finally, the user could use the Auto File button and a popup window would ask which folder to automatically file this and all future emails from this address upon arrival. The user is also offered to create a new folder if the appropriate one does not already exist.

Emails that are deleted in step 4 or 5, or as a result of being placed on a list by steps 4 or 5, may be moved into the Deleted_Spam folder. Going to that folder and using the new Undelete key moves the message to the normal inbox and removes the email address or content from the always delete lists, but this may not return deleted email addresses to the relationship tree.

The techniques disclosed may provide the following advantages in one or more embodiments:

1. Parsing nested email addresses into a social network relationship tree that captures and preserves the multiple levels and interconnections, of email address relationships within a users private email corpus.

2. Use of the data in a social network relationship tree to determine and report the multiple paths of introduction to targeted individuals.

3. Sharing of personal social network relationship tree with others in order to expand the extent of contacts, i.e. the method of creating an extended social network relationship tree.

4. Sharing of personal social network relationship tree with others without disclosing the contents of the relationship tree that are not on direct paths to the target.

5. Use of the personal social network relationship tree in the filtering of undesirable bulk email advertising such as spam.

6. Use of the social network relationship tree to market products to personal contacts, and to their contacts and again to their contacts.

7. The method of building a confidential skills profile compendium that provides access only to individuals that are within a certain "diameter" or "distance", from the individual whose skills are recorded, based upon the inquirers personal and extended social network relationship tree.

8. Use of the "all users" aggregate database to provide an "email scoring" service that identifies email addresses as having historical communications activities that are statistically typical of addresses used for certain purposes, such as fraudulent purposes.

9. Use of the user's relationship tree to find an individual known to the user directly, or through introduction, that has experience with a particular commerce activity at a participating website.

Referring now to Fig. 1, correspondence comes in many forms including printed correspondence delivered by post or forwarded by facsimile, email correspondence as well special purpose correspondence such as telephone bills. Document 11, for example, is a piece of correspondence sent by Tom, the addressor, to Bill, the addressee. Bill and Tom are the primary addressed parties and form a correspondence, or contact pair, at the ends of a contact or correspondence path from Tom to Bill. As shown in document 11, there may be other parties to the correspondence addressed at a different level, such as secondary addressees Jane and John, who are addressed directly in document 11 by being indicated to receive copies of document 11. In particular, Jane and John are each separate direct addressees at the end of a contact path from Tom although they have some level of connection as noted below.

Certain types of correspondence may also include addressed parties not directly addressed, that is indirectly addressed, in the current document. For example, document 11 may be a document forwarding a copy of other correspondence, such as document 13, which includes indirectly addressed parties Jim, George, Mary, Tom and John. Other types of correspondence, such as telephone bills, may include indirectly addressed parties in that information such as each identified telephone number called indicates at least one address form representing an addressed party even though the phone bill is not directed to any of these indirectly addressed parties. Each indirectly addressed party on a telephone bill may be on the end of a contact path from the phone bill's addressee while the primary or direct contact path is from the phone company to the billed addressee.

Referring now to Fig. 2, each addressed party in a piece of correspondence may be said to have a relationship, such as a contact path, with the other directly addressed parties. For example, as shown, Bill and Tom may be said to be the ends of a contact pair as a result of document 11. This contact pair may be identified by contact path 15 from Tom, the addressor, to Bill, the addressee. The direction of the path may be indicated by the direction of the arrowhead or other means on contact path 15. Further, Jane and John are each at the end of a contact path from Tom shown as contact paths 17 and 19, respectively.

Contact paths, in addition to having at least a pair of addressed parties, also at least potentially include additional or secondary information, such as the direction of flow of the correspondence and/or whether or not the parties were directly or indirectly addressed in the document being considered, such as document 11. Additionally this information could include all the dates of communication, pointers identifying the specific communication or the source of communication or any other meaningful information that can be extracted from the original source data. For convenience, contact paths 15, 17, 19, 21, 23, 25 and 27 are shown with arrowheads to indicate the direction of contact In summary, contact paths between addressed parties may therefore include secondary information such as the direction of correspondence as well as the addressed pair of parties. Depending on the intended usage, data

collected with regard to addressed parties may include such secondary information for some types of contact paths and may not include such secondary information for other types of contact paths.

Referring now to Fig. 3, the process will be described in terms of steps taken with regard to a first user, User A, to develop a local data file, and/or the combination of that data with data from a similar user, such User B not shown, to create a web relational data base or database of relationships, followed by descriptions of a series of services or tools that may interact with the database of relationships.

Beginning with User A, step 10 operates to choose a group of email records to process. In step 12, record headers or equivalent text are parsed, including those in nested or forwarded email messages, in order to retrieve email addresses for all addressed parties along with From:, To: and Cc: relationships for each address. Thereafter, in step 14, data may be extracted, or an algorithm may be applied to each email and attachments, that provides a unique numeric result for each email processed as a unique source ID. In step 16, data may be written to a data store such as a local hard drive, for example as a relational or flat file 18, to temporarily store the extracted email headers and relationship information as well as the unique source ID.

Some of the functions may then be performed locally for User A based on data collected in flat file 18, but substantial advantages can be achieved by subsequent processing to create a Internet based relational database such as central web based UDDI relational data base 20. A UDDI, or Universal Discover, Description and Integration database, is a standards based XML database with restricted or controlled access to the data. In particular, in step 22, data is uploaded to a central web based relational database 20 which is protected by user ID and password available only to the user. In step 24, the user may optionally designate other users that have permission to access the owner's data.

The data to be written to relational data base 20 may then be processed by server side database pre-processing operations in step 40 with filters that prevent duplicates and process only incremental data from the flat file. Step 40 may also key data to the user providing that data so it is only accessible by authorized users which may have been designated in step 24. Step 40, in addition to uploading the preprocessed data to relational database 20, may also cause the writing back of data to local files, such as data file 18, to facilitate further processing by reducing need to reprocess previously processed data.

Once the relevant data has been uploaded to relational database 20, which may conveniently be accessible to a group of users by for example being located on a central server in a local network or preferably in a wide area network such as the Internet, various processes or tools may be used to work with this data.

Referring now in more detail also to Figs. 4 and 5, relationship visualization tool 44 may provide visualization by display for the user of contact relationship data in central database 20 by loading the data in step 46 that the user is authorized to access. In step 48, data points representing contacts or addressed parties may be arranged to identify the most frequent links. Color codes, based upon recency of contact and/or degrees of separation, may be assigned. The spatially arranged and color coded results may then be displayed on display monitor 50. The results displayed on monitor 50 may represent the relationships, and paths there between, beginning with the user and extending through all contacts, or addressed parties, disclosed in the emails, or other source of data, processed by the steps disclosed and may be referred to herein as a relationship tree which shows the direct and indirect relationships of a user.

As shown in Fig. 5, the data visualized from database 20 may show, for example, that User A has direct relationships, at least with regard to one or more existing emails, with Contacts B and E, while Contact B has additional direct relationships with Contacts C and F while Contact C has a direct relationship with Contact D and Contact E has a direct relationship with Contact F. Although a typical useful

visualization display of this type may be much more complicated than as shown in Fig. 5, it is apparent that User A may much more easily comprehend that he can make contact with Contact D via Contacts B and C by viewing the visualization in Fig. 5 than be reading the above provided text.

Referring now in greater detail to Figs. 6 and 7, referral path identification 52 operates on the data, in step 54, by loading data that the user is authorized to access. The user may then input target email address(s), or any other valid search criteria such as that available from directories cross referenced to email addresses, in step 56. The data and email address(s) may then be processed in step 58 using a breadth-wise incremental search to determine linkage paths which are then used to create display 60 in which the results may be displayed as highlighted paths or list of contacts.

As shown in Fig. 7, the closest path between User A and Contact D, the inputted email address, is shown as the highlighted path via Contacts B and F. It should be noted that a similar length path happens to exist via Contacts E and F, but is not shown as highlighted. The selection of the path via Contacts B and F may be made automatically in processing step 58 on the basis of the most recent contacts made along this path of parts of it, on the basis of the number of contacts made along this path of parts of it and preferably upon a combination of both the above described recency and frequency criteria.

Spam filter 62 may operate upon data provided by the user in step 64 indicating the degrees of freedom or separation, the to use as a filter on the data loaded in step 66. A single degree of freedom or a single step of separation refers to a direct contact, such as the relationship between User A and Contact B in Fig. 5. A second degree of freedom, or two steps of separation, refers to the indirect relationship between User A and Contacts C and D in Fig. 5.

In step 68, inbound emails with origination addresses that match relationship tree addresses in accordance with the degrees of freedom data provided in step 64 are

placed in a filtered inbox. Inbound emails with origination addresses not matching addresses on the relationship tree may be left in the general inbox for review or may be further filtered based on other criteria to evaluate the likelihood that they are undesired emails such as SPAM.

As shown in Fig. 10, multilevel marketing (MLM) & referral marketing step 70 combines the degrees of separation selection provided by the user in step 72, and a marketing offer or other letter provided by the user in step 74, with data loaded in step 76 to personalize each letter with the referrer's email address in merge program 78.

Referring now to Fig. 10, skill registry tool 80 may be used to obtain introductions to individuals with specific skills. The user provides a selected degree of separation in step 82 together with data related to the desired skill set, and/or experience, in step 84 which are compared with the relationship tree lists to form a qualified email list 86. List 86 may be further processed in step 88 with a breadth-wise incremental search to determine linkage paths for creating display 90 which may display results as highlighted paths or list of contacts. Other directories may be cross referenced to provide expanded search capabilities.

Additionally online registry 92 may be made available for individuals to post answers to detailed questions about their skills and experience while providing an email address. Data from online registry 92 may then be loaded from database 20 in step 94 and added for processing in list 86 to further qualify the email lists.

Referring now to Fig. 11, interface 96 may be used to provide and monitor licensed access to data in step 98 in which data is made available to third party software providers who can develop products that utilize the relationship tree database. Access to the data remains restricted to the owners of the data.

Referring now to Fig. 12, interface 100 may be used to provide a user with a reference from an individual known to the user regarding commerce activities at a

participating website. Typically a context sensitive link 102 allows the user to expose their relationship tree 108, and the website to expose a visitor history file 112 from patrons who have elected to participate at 110. The data is then matched for relevance in step 104 and then filtered data is made available to the user in step 106, where a list of potential endorsers is made known, possible with their posted comments.

Referring now to Fig. 13, interface 120 may be used to provide credit issuers (or credit card sales retailers) an additional means of evaluating the credit worthiness of a particular transaction. Proprietary algorithms are employed at 122 to periodically review the pattern of connections of all email addresses in the database. This is performed on communication link history from all relationship trees without regard to the owners of the information. The algorithm assigns a "score" that indicates a deviation from "normal" usage. Authorized subscribers can make inquiries at 124 that reveal the "score" at 126. Authorized subscribers use this information along with other information they already have to help them in their decision regarding the validity of the transaction.

I Claim:

1. A method for developing contact information from correspondence such as emails, comprising:

processing a set of correspondence to develop a database of relationships between directly and indirectly addressed parties provided by one or more users.

maintaining the database, on a network in which addressed parties from a plurality of users are combined, by further processing later received correspondence; and

utilizing the database of relationships to provide relationship information between at least one of said users and the addressed parties.

2. The method of claim 1 further comprising:

associating a unique identification with each piece of correspondence; and

using the unique identification to detect duplications of correspondence in order to more accurately determine a frequency of communication between addressed parties.

3. The methods of claims 1 or 2 further comprising:

displaying connection paths between each of said users and at least some of the addressed parties.

4. The method of claim 3 comprising:

displaying additional addressed par +ties upon selection of certain displayed addressed parties.

5. The method of claim 3 further comprising:

visually identifying intermediate addressed parties, if any, between said one of said users and a selected addressed party.

6. The method of claim 3 further comprising:

visually identifying the frequency of correspondence in the database of relationships between said one of said users and at least some of said addressed parties.

7. The method of claim 3 further comprising:

visually identifying the most recent correspondence between at least some of said addressed parties.

8. The method of claim 7 further comprising:

visually identifying the frequency of correspondence in the database of relationships between said one of said users and at least some of said addressed parties.

9. The method of claim 3 wherein the connection paths displayed are prioritized in accordance with the number of intermediate addressed parties, if any, between said one of said users and a selected addressed party.

10. The method of claim 3 wherein the connection paths displayed are prioritized in accordance with the most recent correspondence between said one of said users and said selected addressed party.

11. The method of claim 10 wherein the connection paths displayed are prioritized in accordance with the frequency of correspondence between said user and said pre-selected addressed party.

12. The methods of claims 1, 2, 9, 10 or 11 further comprising:

sorting incoming correspondence in accordance with the number of intermediate contacts, if any, identified in the database of relationships between said user and the addressors of said incoming correspondence.

13. The method of claims 1, 2, 9, 10 or 11 further comprising:

addressing outgoing correspondence to parties selected in accordance with the number of intermediate contacts, if any, identified in the database of relationships between said user and said addressed parties.

14. The method of claims 1, 2, 9, 10 or 11 further comprising:

combining said database of relationships with data related to the skills and experience of third parties to identify paths between said user and third parties having selected skills and experience.

15. The method of claims 1, 2, 9, 10 or 11 further comprising:

combining said database of relationships with data related to the shopping experiences of third parties to identify paths between said user and third parties having selected shopping experiences.

16. The method of claims 1, 2, 9, 10 or 11 further comprising:

analyzing said database of relationships in accordance with statistic norms to determine deviations from such statistical norms in patterns of correspondence of a selected addressed party.

17. A method for deriving qualitative information related to addressed parties on correspondence such as emails, comprising:

processing a set of correspondence to develop a database of relationships between addressed parties;

maintaining the database by further processing later received correspondence; and

utilizing the database of relationships to determine patterns of correspondence for addressed parties.

18. The method of claim 17, wherein the processing further comprises:

processing indirectly addressed parties to develop the database of relationships between directly and indirectly addressed parties.

19. The method of claims 17 or 18 further comprising:

associating a unique identification with each piece of correspondence; and

using the unique identification to detect duplications of correspondence in order to more accurately determine a frequency of communication between addressed parties.

20. The method of claims 17 or 18 further comprising:

maintaining the database of relationships on a web based database of relationships in which addressed parties from correspondence of a plurality of users are combined.

21. The method of claims 17 or 18, further comprising:

identifying a frequency of correspondence in the database of relationships between a selected addressed party and other addressed parties.

22. The method of claim 21 further comprising:

identifying the most recent correspondence between said selected addressed party and other addressed parties.

23. The method of claims 17 or 18 further comprising:

deriving normal patterns of correspondence between addressed parties; and

determining if patterns of correspondence for the selected addressed party is consistent with the derived normal patterns.

24. A method for developing contact information from correspondence such as emails, comprising:

processing a collection of a user's correspondence to develop a database of relationships between said user and parties directly and indirectly addressed in said correspondence;

maintaining the database by further processing later received correspondence; and

utilizing the database of relationships to provide relationship information between said user and the addressed parties.

25. The method of claim 24 further comprising:

associating a unique identification with each piece of correspondence; and

using the unique identification to detect duplications of correspondence before maintaining the database in order to more accurately determine a frequency of communication between said user and the addressed parties.

26. The method of claim 24 further comprising:

maintaining the database of relationships on a web based database of relationships in which addressed parties from other sources may be combined.

27. The methods of claims 24, 25 or 26 further comprising:

displaying connection paths between said user and at least some of the addressed parties.

28. The method of claim 27 comprising:

displaying additional addressed parties upon selection of certain displayed addressed parties.

29. The method of claim 27 further comprising:

displaying intermediate addressed parties, if any, between said user and a selected addressed party.

30. The method of claim 27 further comprising:

displaying a frequency of correspondence in the database of relationships between said user and selected addressed parties.

31. The method of claim 27 further comprising:

displaying a most recent correspondence between the user and selected addressed parties.

32. The method of claim 31 further comprising:

display a frequency of correspondence in the database of relationships between said user and selected addressed parties.

33. The method of claim 27 wherein the connection paths displayed are prioritized in accordance with the number of intermediate addressed parties, if any, between said user and one or more addressed parties.

34. The method of claim 27 wherein the connection paths are prioritized in accordance with the most recent correspondence with selected addressed parties.

35. The method of claim 34 wherein the connection paths displayed are prioritized in accordance with the frequency of correspondence between said user and said pre-selected addressed party.

36. The methods of claims 33, 34 or 35 further comprising:

sorting incoming correspondence in accordance with the number of intermediate contacts, if any, identified in the database of relationships between said user and the addressors of said incoming correspondence.

37. The method of claims 33, 34 or 35 further comprising:

addressing outgoing correspondence to parties selected in accordance with the number of intermediate contacts, if any, identified in the database of relationships between said user and selected addressed parties.

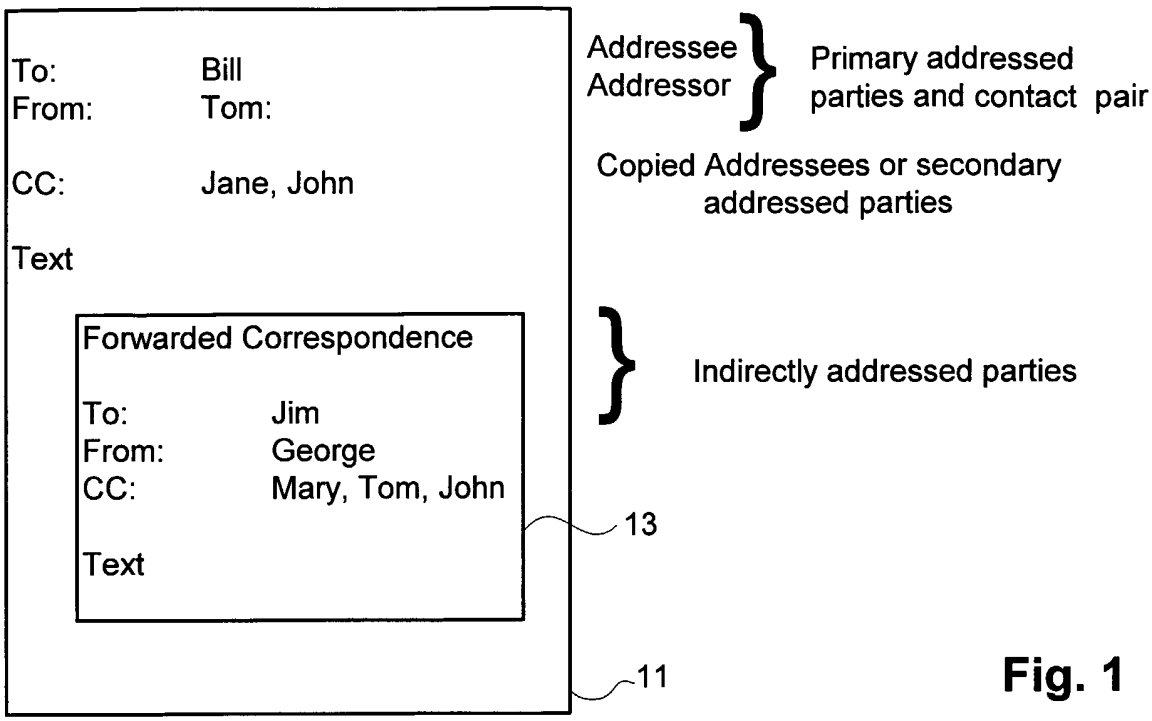

To: Bill
From: Tom:
CC: Jane, John
Text
Forwarded Correspondence
To: Jim
From: George
CC: Mary, Tom, John
Text
13
11
Addressee
Addressor
Primary addressed parties and contact pair
Copied Addressees or secondary addressed parties
Indirectly addressed parties

Fig. 1

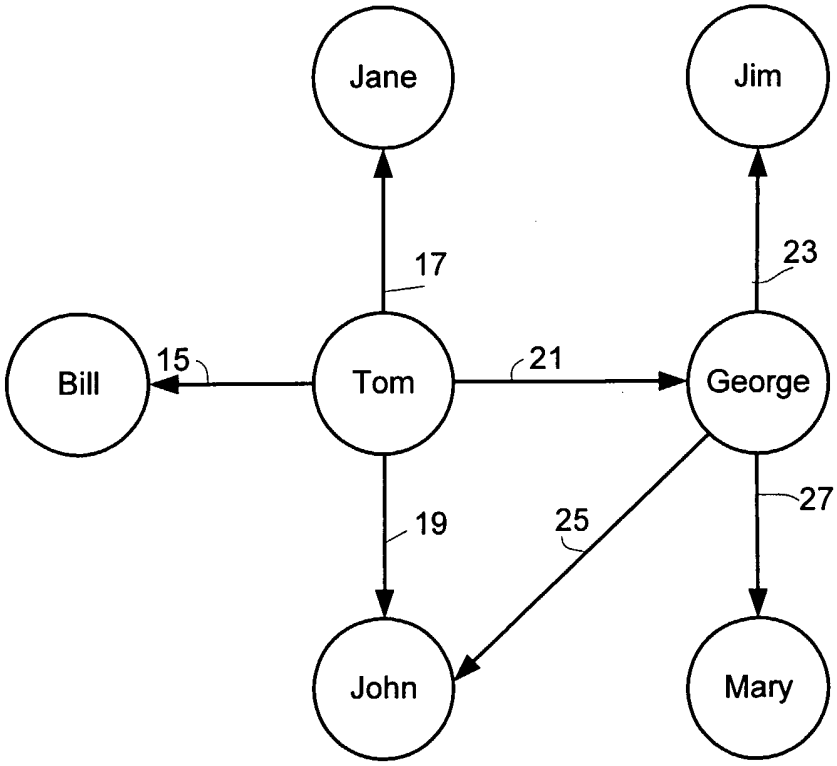

Jane
Jim
17
23
Bill
15
Tom
21
George
19
25
27
John
Mary

Fig. 2

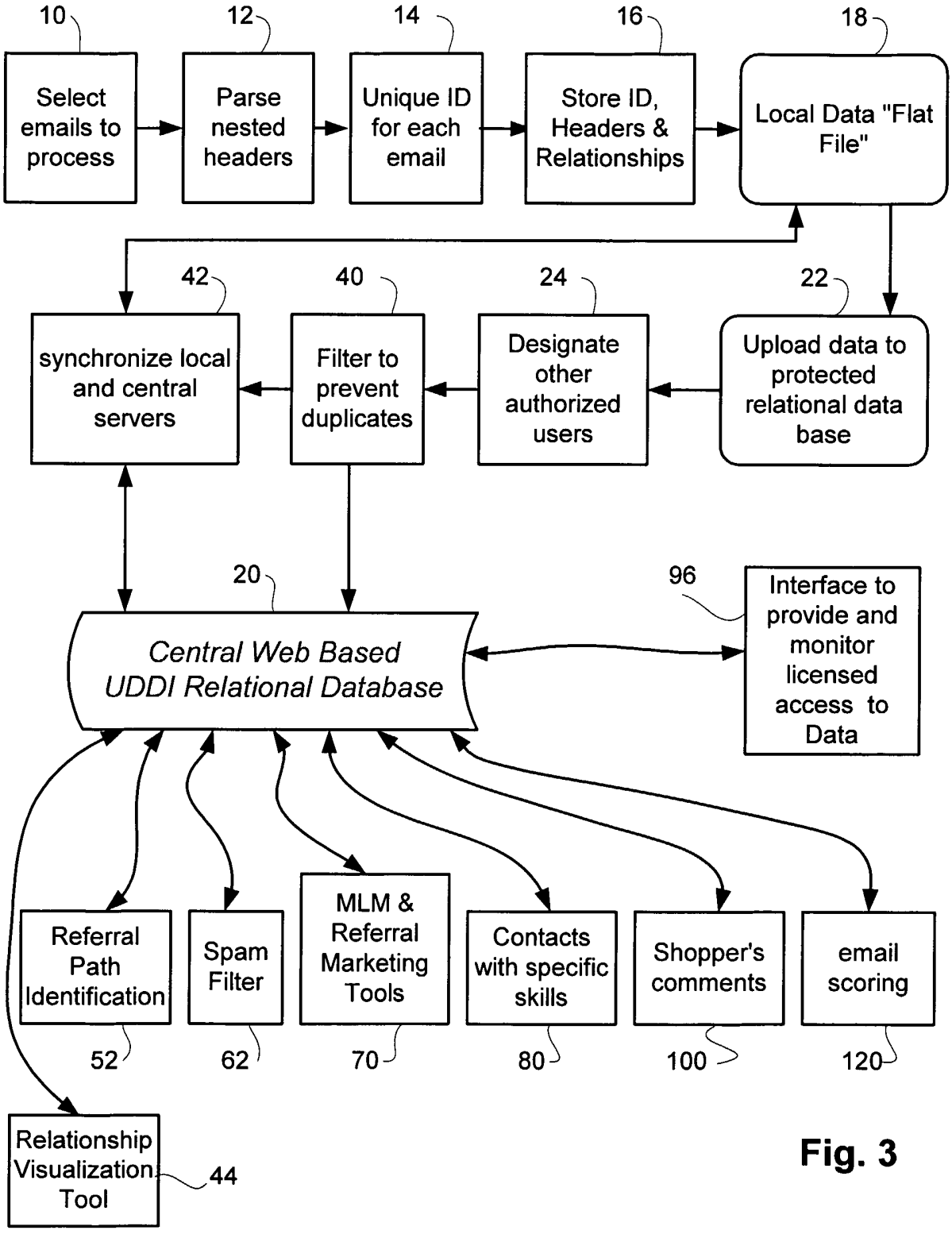

10
Select emails to process
12
Parse nested headers
14
Unique ID for each email
16
Store ID, Headers & Relationships
18
Local Data "Flat File"
42
synchronize local and central servers
40
Filter to prevent duplicates
24
Designate other authorized users
22
Upload data to protected relational data base
20
Central Web Based UDDI Relational Database
96
Interface to provide and monitor licensed access to Data
Referral Path Identification
52
Spam Filter
62
MLM & Referral Marketing Tools
70
Contacts with specific skills
80
Shopper's comments
100
email scoring
120
Relationship Visualization Tool
44

Fig. 3

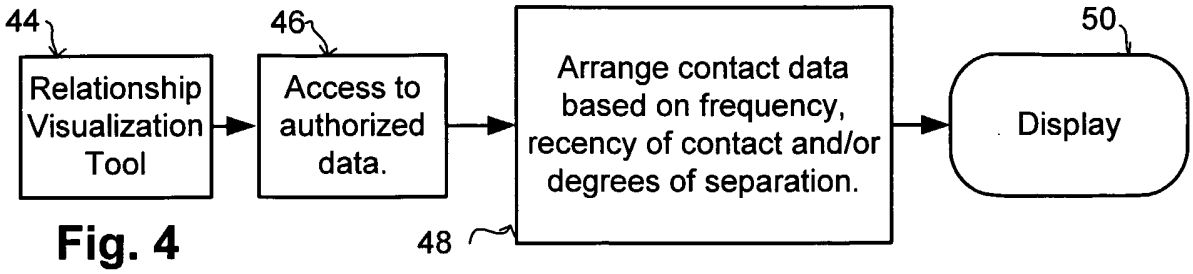


Fig. 4

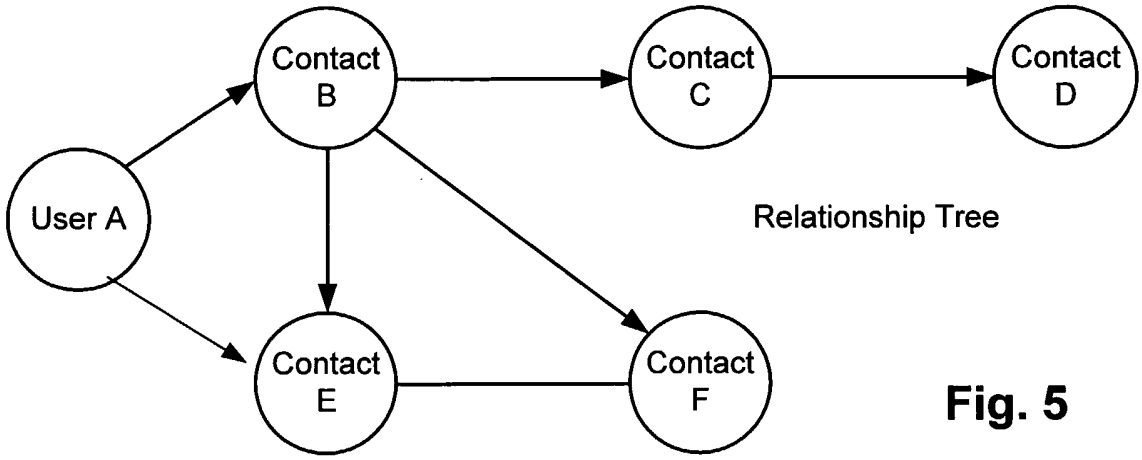


Fig. 5

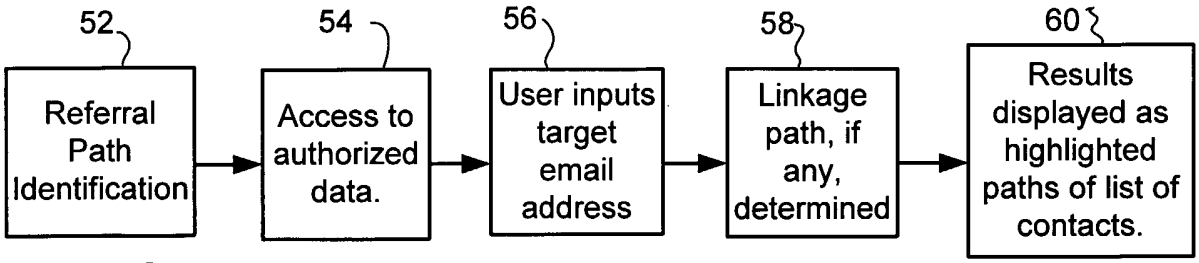


Fig. 6

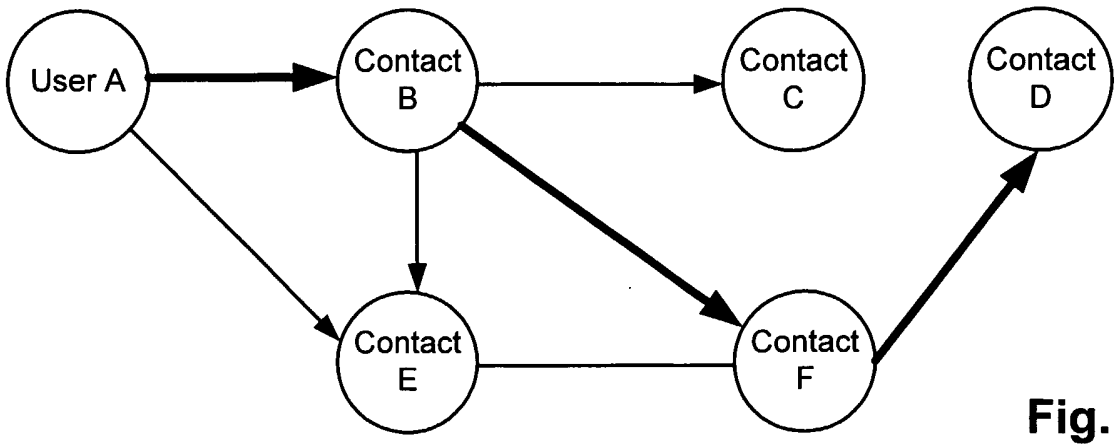


Fig. 7

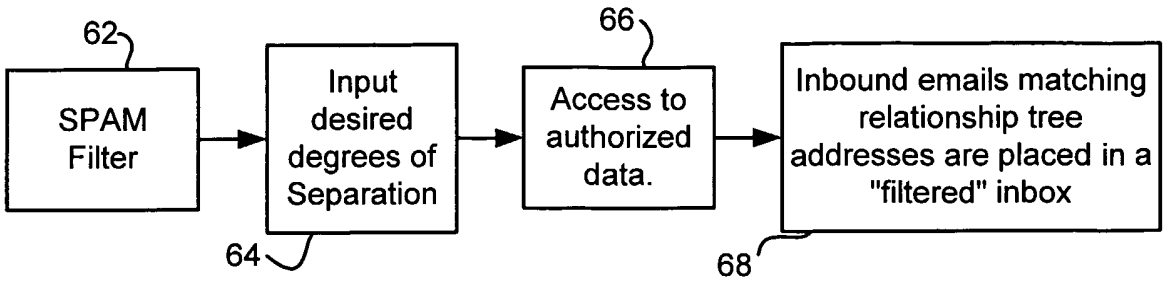


Fig. 8

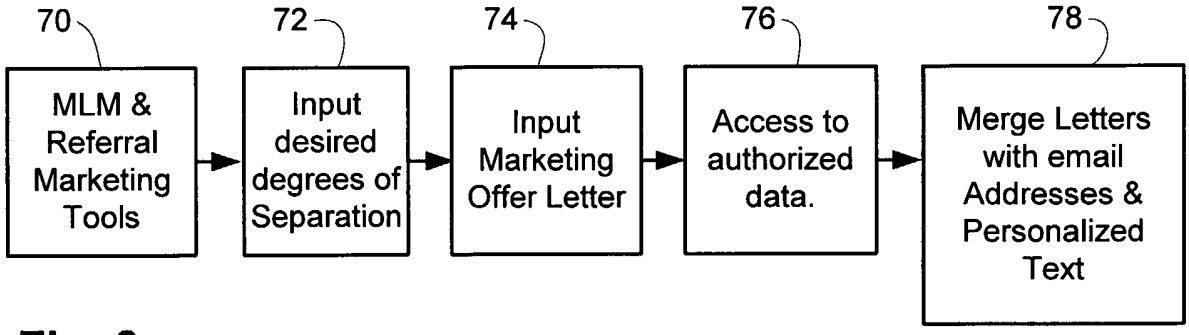


Fig. 9

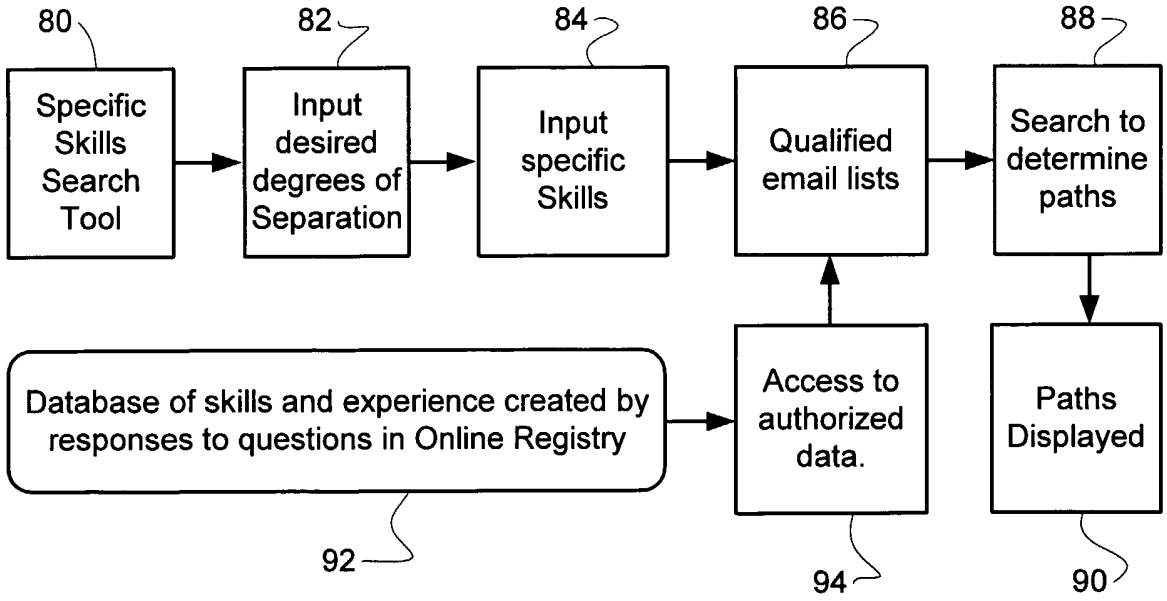


Fig. 10

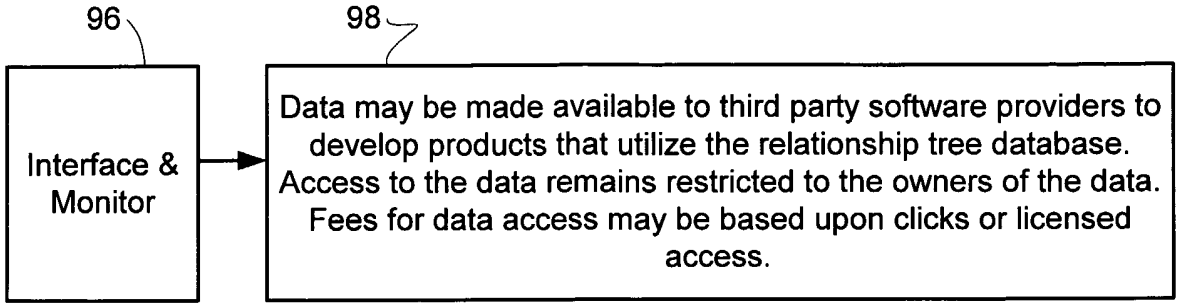


Fig. 11

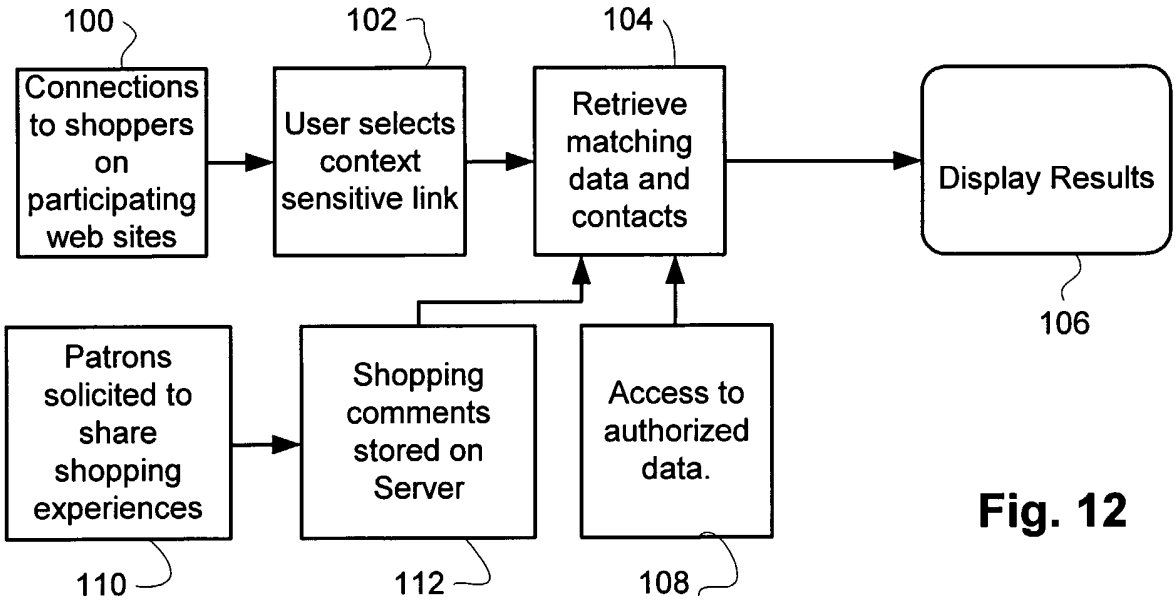


Fig. 12

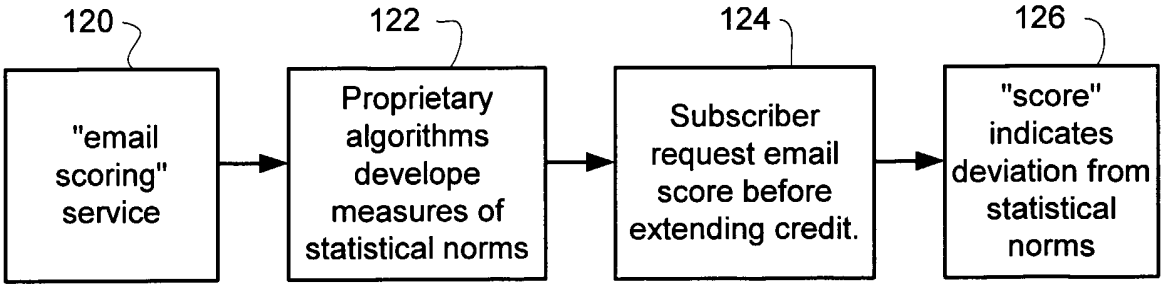


Fig. 13